

# Entropic nonsignaling correlations

Rafael Chaves<sup>1,2,3</sup> and Costantino Budroni<sup>4</sup>

<sup>1</sup>*Institute for Physics & FDM, University of Freiburg, 79104 Freiburg, Germany*

<sup>2</sup>*Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany*

<sup>3</sup>*International Institute of Physics, Universidade Federal do Rio Grande do Norte, 59070-405 Natal-RN, Brazil*

<sup>4</sup>*Naturwissenschaftlich-Technische Fakultät, Universität Siegen, Walter-Flex-Straße 3, 57068 Siegen, Germany*  
(Dated: June 16, 2016)

We introduce the concept of entropic nonsignaling correlations, i.e., entropies arising from probabilistic theories that are compatible with the fact that we cannot transmit information instantaneously. We characterize and show the relevance of these entropic correlations in a variety of different scenarios, ranging from typical Bell experiments to more refined descriptions such as bilocality and information causality. In particular, we apply the framework to derive the first entropic inequality testing genuine tripartite nonlocality in quantum systems of arbitrary dimension and also prove the first known monogamy relation for entropic Bell inequalities. Further, within the context of complex Bell networks, we show that entropic nonlocal correlations can be activated.

Quantum nonlocality—the fact that correlations obtained in quantum experiments performed by distant parties are incompatible with local hidden variable (LHV) models [1]—brings to light an intriguing aspect of quantum mechanics (QM) and relativistic causality [2]. QM is in accordance with the nonsignaling (NS) principle, that is, local manipulations by an experimenter cannot influence the measurement outcomes of other distant experimenters. However, as demonstrated by Popescu and Rohrlich [3], special relativity alone cannot single out quantum mechanical correlations as there are theories, beyond QM, also in agreement with NS. This result not only has triggered the search for physically well-motivated principles for quantum mechanics [4–9] but also has led to new insights about its limitations for information processing [10–14].

Given the intrinsic statistical nature of QM, probabilities give a natural framework for nonlocality. Indeed, Bell inequalities and NS relations are nothing other than constraints on probabilities arising in a given theory, local and NS, respectively [15]. Nevertheless, different approaches are possible [16–18]. In particular, in the information-theoretic approach to nonlocality [17–23] the basic objects are the Shannon entropies [24] of the observed data.

The information-theoretic approach provides a novel and useful alternative for both conceptual and technical reasons. First, entropy is a key concept in both classical and quantum information theory, thus developing a framework that focuses on entropies rather than probabilities leads to new insights and applications [25–32]. For instance, the celebrated principle of information causality [6] is nothing other than an entropic inequality bounding the correlations that can be achieved by imposing a certain causal structure to quantum mechanics [14]. Second, entropies allow for a much simpler and compact characterization of classical and quantum correlations in a variety of scenarios

[14, 21, 29, 33]. In spite of that, as opposed to the usual probabilistic description, little is known about entropic Bell inequalities beyond very simple cases and remarkably nothing is known about the structure imposed by the nonsignaling principle on the entropies of measurement outcomes.

In this letter, we aim to further develop the information-theoretic approach to nonlocality and, in particular, to define the concept of entropic nonsignaling correlations, i.e., the entropies compatible with the nonsignaling principle. We characterize NS entropic correlations in a variety of scenarios: from usual bipartite and tripartite, to genuine multipartite nonlocality [34–36], bilocality [37], and information causality [6]. Our framework can also be employed to derive monogamy relations [38, 39] between entropic Bell inequalities. Furthermore, our methods highlight the use of entropic NS correlations as a novel tool to derive Bell inequalities in scenarios otherwise intractable.

*Marginal scenarios, local and NS correlations.*— In a quantum experiment, only some of the relevant observables are jointly measurable; hence, we face fundamental restrictions on the empirically accessible joint probability distributions. This fact is encoded in the notion of a marginal scenario. Given  $n$  random variables  $\{X_1, \dots, X_n\}$ , a marginal scenario  $\mathcal{M}$  is defined as  $\mathcal{M} = \{S_1, \dots, S_{|\mathcal{M}|}\}$ ,  $S_i \subseteq \{X_1, \dots, X_n\}$ , such that for each  $S_i$  a joint probability distribution  $P[(X_s)_{s \in S_i}]$  is accessible [18, 20]. Clearly, it is sufficient to consider maximal subsets.

A typical example is a Bell experiment: two separated parties, Alice and Bob, at each run of the experiment can perform one of  $m$  different measurements, labeled as  $A_x$  and  $B_y$ , respectively, on their shares of a joint system. Their marginal scenario is, then,  $\mathcal{M}_{\text{Bell}} = \{\{A_x, B_y\}\}_{x,y=1,\dots,m}$ , corresponding to the probability distributions  $\mathbf{p}_{\text{obs}} = p(a_x, b_y)$  [40]—where  $a_x$  labels

the outcome when measurement  $x$  has been performed (similarly for  $b$ )—estimated from the statistical data. As shown by Fine [41], a LHV model for the data can be equivalently defined as a joint probability distribution  $\mathbf{p} = p(a_1, \dots, a_m, b_1, \dots, b_m)$ . Hence, a set of marginals is called local if it is consistent with a single joint probability distribution for all measurements. This, in turn, implies the existence of a joint entropy of all possible measurements  $H_{A_1 \dots A_m B_1 \dots B_m}$  and all its marginals [17], where  $H_X := H(X) := -\sum_x p(x) \log_2 p(x)$  stands for the Shannon entropy. They can be represented as a  $2^{2m}$ -dimensional vector  $\mathbf{h} = (H_{\emptyset}, H_{A_1}, \dots, H_{A_1 \dots A_m B_1 \dots B_m})$ . Notice that  $H_{\emptyset}$  is defined to be 0, but it is convenient to include it to have a more compact representation of the constraints satisfied by the entropy vector (cf. Appendix IA).

The difference between the probabilistic and entropic description solely relies on how we quantify correlations. A marginal probability distribution  $\mathbf{p}_{\text{obs}}$  is local if we can construct a well-defined joint probability distribution  $\mathbf{p}$ . Similarly, marginal entropies  $\mathbf{h}_{\text{obs}} = H(A_x, B_z)$  are local if a joint entropy and all its marginals  $\mathbf{h}$ , arising from a (nonunique) joint probability distribution, can be defined. The existence of a well-defined joint description  $\mathbf{p}$  imposes strict constraints—the famous Bell inequalities—on the empirically observable marginal correlations  $\mathbf{p}_{\text{obs}}$  [41–43]. Similarly, marginal entropic correlations  $\mathbf{h}_{\text{obs}}$  admitting an extension to  $\mathbf{h}$  also obey strict constraints. The closure set of well-defined entropy vectors  $\mathbf{h}$  defines a convex cone  $\Gamma_E$ , that is, if  $\mathbf{h}$  and  $\mathbf{h}'$  are in  $\Gamma_E$  so are  $p\mathbf{h} + (1-p)\mathbf{h}'$ , with  $0 \leq p \leq 1$ , and  $\lambda\mathbf{h}$ , with  $\lambda \geq 0$  [44]. An explicit characterization of  $\Gamma_E$  is yet to be found, however, an outer approximation, characterized by finitely many linear inequalities [24] or, equivalently, in terms of finitely many extremal rays (vectors defined up to a positive factor [45]), is known: the Shannon cone  $\Gamma_{\text{Sh}}$ . Such inequalities basically amount to the positivity of the conditional entropy, i.e.  $H(A|B) := H(A, B) - H(B) \geq 0$ , and the positivity of the conditional mutual information, i.e.,  $I(A : B|C) := H(A, C) + H(B, C) - H(A, B, C) - H(C) \geq 0$ , for disjoint subsets of variables  $A, B, C$  (see Appendix IA for further details). Thus, in full analogy with the probabilistic case [46], entropic Bell inequalities can be understood as the constraints arising from the projection of  $\Gamma_E$  onto observable coordinates, that is, the projection of  $\mathbf{h}$  into  $\mathbf{h}_{\text{obs}}$  defining the Bell entropic cone  $\Gamma_{\text{Bell}}$ .

On the other hand, NS probabilities are defined as those where the outcomes of a part do not depend on the measurements performed by another distant part, i.e., such that  $p(a_x) = \sum_{b_y} p(a_x, b_y) = \sum_{b_{y'}} p(a_x, b_{y'})$  (similarly for  $b$  and for any number of parties). NS correlations are then defined by the above linear constraints (NS conditions) together with the nonnegativ-

ity condition  $p \geq 0$ ; i.e., they are classical probability distributions whenever restricted to  $p(a_x, b_y)$ , with consistent marginals. Geometrically, they can be seen as the intersection of the simplex polytopes [47] defining each of the probabilities  $p(a_x, b_y)$  and thus overlapping over the marginals  $p(a_x)$  and  $p(b_y)$ . We can then naturally define NS entropic cone, for a marginal scenario  $\mathcal{M} = \{S_1, \dots, S_{|\mathcal{M}|}\}$ , as the intersection  $\Gamma_{\text{NS}} = \Gamma_i \cap \dots \cap \Gamma_{|\mathcal{M}|}$ , where  $\Gamma_i$  is the entropy cone associated with  $S_i$  (see Fig. 1). For instance, in the bipartite scenario, the NS cone is given by the intersection of  $2m$  cones corresponding to the subsets of variables appearing in the marginal scenario  $\mathcal{M}_{\text{Bell}} = \{\{A_x, B_y\}\}_{x,y=1,\dots,m}$  and respecting the basic constraints given by  $H(A_x|B_y) \geq 0$ ,  $H(B_y|A_x) \geq 0$  and  $I(A_x : B_y) \geq 0$ . This intersection can be understood as follows: since each  $S_i$  contains a restricted set of variables, we embed each  $\Gamma_i$  in a bigger space where the variables not in  $S_i$  are unconstrained.

In the following, we apply the above framework to analyze from an entropic perspective a broad range of scenarios. Notice that for  $n \leq 3$  variables, the entropy cone corresponds to the Shannon cone, i.e.,  $\Gamma_E^n = \Gamma_{\text{Sh}}^n$  [24]; hence, all results for the bipartite and tripartite cases lead to the exact description of the NS cones [48]. Further discussions and technical details can be found in [49], including the derivation of all Bell inequalities that, nicely, can be proven by simple sums of Shannon inequalities.

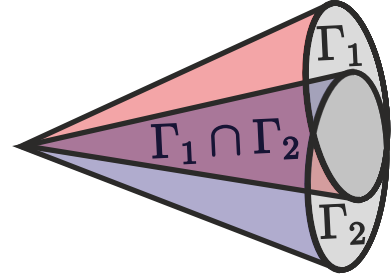


FIG. 1. Pictorial illustration of the NS entropic cone.

*Bipartite and tripartite scenarios.*— We start with the simplest Bell scenario as above, for  $m = 2$ . In contrast to the probabilistic case, entropic correlations are concisely defined for an arbitrary number of measurement outcomes, highlighting another advantage of the entropic approach. For dichotomic observables ( $a_x, b_y = 0, 1$ ), the only nontrivial probabilistic Bell inequality is the Clauser-Horne-Shimony-Holt (CHSH) inequality [50]

$$S = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle - 2 \leq 0, \quad (1)$$

where  $\langle A_x B_y \rangle = \sum (-1)^{a_x + b_y} p(a_x, b_y)$  stand for the ex-

pectation value. Its entropic version [17, 18],

$$S_E = I_{A_0:B_0} + I_{A_0:B_1} + I_{A_1:B_0} - I_{A_1:B_1} - H_{A_0} - H_{B_0} \leq 0, \quad (2)$$

is valid for any number of outcomes, where  $I_{A_x:B_y} = H_{A_x} + H_{B_y} - H_{A_x B_y}$  represents the mutual information.

Both inequalities are maximally violated by an extremal point or ray characterizing the NS correlations. Eq. (1) is maximally violated by the Popescu-Rohrlich (PR)-box  $p_{\text{PR}}(a_x, b_y) = (1/2)\delta_{a \oplus b, xy}$ . However,  $p_{\text{PR}}$  is entropically equivalent to the classical correlation  $p_{\text{C}}(a_x, b_y) = (1/2)\delta_{a \oplus b, 0}$  and thus cannot violate (2). On the other hand, Eq. (2) is maximally violated by  $H(A_x, B_y) = \log_2(d)(1 + xy)$ , with marginals  $H(A_x) = H(B_y) = \log_2(d)$  and  $d$  the number of outcomes. Thus, this correlation can be interpreted as the entropic counterpart of a PR-box. For  $d = 2$ , these entropies are obtained as an equal mixture of  $p_{\text{PR}}$  and  $p_{\text{C}}$ . The mixing with  $p_{\text{C}}$  is exactly the method proposed in [51] to turn entropic inequalities into necessary and sufficient conditions for nonlocality detection. It is thus appealing that the NS entropic cone naturally retrieves this sort of correlations.

Another important result of our approach is the derivation of the first entropic monogamy relation for Bell inequalities. The monogamy of Bell inequalities violations is a general feature of NS theories [38], and it can be understood by the following example. For a tripartite distribution  $p(a_x, b_y, c_z)$  with binary inputs and outputs, whenever the marginal distribution  $p(a_x, b_y)$  violates the CHSH inequality necessarily  $p(a_x, c_z)$  must be local. Similarly, from the definition of NS entropic cone, we can prove that

$$S_E^{AB} + S_E^{AC} \leq 0, \quad (3)$$

meaning that both entropic Bell inequalities, between Alice-Bob and Alice-Charlie, cannot be violated at the same time, with the notable difference that this monogamy inequality is valid for any number of outcomes.

The similarities between the probabilistic and entropic approaches, which may suggest a deeper geometric connection [21], already disappear in the tripartite scenario. For the case of three parties and two settings, the probabilistic NS correlations for dichotomic measurements consist of 46 different classes of extremal points, with 45 of them nonlocal [52]. In turn, the entropic NS cone is characterized by 1292 different classes of extremal rays, 1164 of which correspond to nonlocal correlations [53]. As it turns out, already at the tripartite case we obtain a much more complex structure than the one we could naively presume from the probabilistic description.

*Genuine tripartite entropic nonlocality.*— In analogy to entanglement [54], when moving beyond the bipartite

case, different classes of nonlocality arise. With three parties, one can introduce the notion of genuine tripartite nonlocality, that is, a stronger form of nonlocality that cannot be reproduced even if any two of the parties are allowed to share some nonlocal resources [34–36]. We focus our attention to nonsignaling resources (e.g., a PR-box) and two possible measurements per party, extensions to more measurements and parties are straightforward. For a given bipartition, say  $A|BC$ , a hybrid local-nonsignaling (L|NS) model is equivalent to the existence of probability distributions  $p(a_0, a_1, b_j, c_k)$ , with consistent marginal  $p(a_0, a_1, b_j)$ ,  $p(a_0, a_1, c_k)$ , i.e., Alice has local correlations and Bob and Charlie share nonsignaling correlations. Genuine tripartite nonlocal (GTNL) correlations correspond to marginals  $p(a_i, b_j, c_k)$  that cannot be explained as a convex combination of models of the type  $A|BC$ ,  $B|AC$ , and  $C|AB$ .

Analogously to the NS case, an entropic  $A|BC$  model corresponds to the joint entropies  $H(A_0, A_1, B_j, C_k)$ ,  $j, k = 0, 1$ , and all its marginals, and similarly for  $B|AC$ , and  $C|AB$ . We can then define the L|NS entropic correlations via the cone  $\Gamma_{\text{L|NS}}$ , constructed as the convex hull (i.e., set of convex combinations) of the entropic cones for each of the models  $A|BC$ ,  $B|AC$ , and  $C|AB$ . In turn, GTNL entropic correlations are those lying outside  $\Gamma_{\text{L|NS}}$ .

From the 1164 different classes of extremal nonlocal rays defining the tripartite scenario, 932 correspond to GTNL correlations. One of these rays correspond to the distribution  $(1/2)(p_{\text{XYZ}} + p_{\text{C}})$ , that is, the mixing of  $p_{\text{XYZ}}(a, b, c|x, y, z) = (1/4)\delta_{a \oplus b \oplus c, xyz}$  [55] with classical correlations  $p_{\text{C}}(a, b, c|x, y, z) = (1/4)\delta_{a \oplus b \oplus c, 0}$ . The GTNL character of this correlation can be witnessed by the violation of the following entropic inequality valid for any L|NS correlation with arbitrary number of outcomes:

$$S_{\text{L|NS}} = H_{A_1 B_1 C_0} + H_{A_1 B_0 C_0} + H_{A_1 B_0 C_1} + H_{A_0 B_1 C_0} - H_{A_1 B_1 C_1} - H_{A_1 B_0} - H_{A_1 C_0} - H_{A_0 C_1} - H_{B_1 C_0} \geq 0. \quad (4)$$

Furthermore, inequality (4) can also be used to witness the GTNL in quantum states, for instance using  $d$ -dimensional Greenberger-Horne-Zeilinger (GHZ) states  $|\text{GHZ}\rangle = (1/\sqrt{d}) \sum_{j=0}^{d-1} |jjj\rangle$  and projective measurements [56]. Results are plotted in Fig. 2 up to  $d = 40$ .

*Activating entropic nonlocality in networks.*— The tripartite scenario permits also another possibility: that the correlations between the parties are mediated by independent sources. The paradigmatic example is the entanglement swapping experiment [57]. Two independent pairs of entangled particles are distributed among three spatially separated parties: Bob receives one particle of each pair, and Alice and Charlie the remaining two. By jointly measuring his particles, Bob can gener-

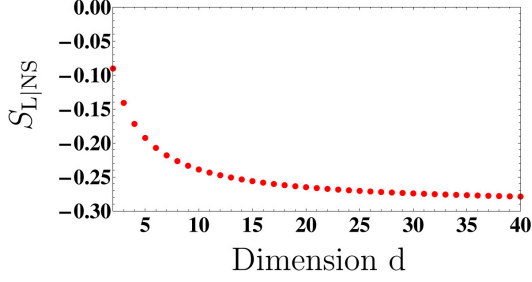


FIG. 2. The violation of inequality  $S_{L|NS} \geq 0$  using GHZ states. The points stand for the violation obtained via numerical optimization.

ate (upon conditioning on outcomes) entanglement and nonlocal correlations between the two remaining particles, even though the latter have never interacted. A probabilistic and local realistic description of this experiment involves two independent hidden variables, the so called bilocality assumption [37, 58–61], implying the independence relation  $p(a, c) = p(a)p(c)$ , i.e., no correlations between Alice and Charlie. The local and NS correlations in the bilocality scenario are defined by infinitely many extremal points and is extremely challenging to characterize [60].

The advantages of the entropic description are here apparent: independence constraints are encoded in simple linear relations, e.g.,  $p(a, c) = p(a)p(c) \rightarrow I(A : C) = 0$ . Geometrically, a set of extra linear constraints  $Lh = 0$ , as the one above, corresponds to the intersection of the (polyhedral) convex cone (e.g.,  $\Gamma_{\text{Bell}}$  or  $\Gamma_{\text{NS}}$ ) with a linear subspace, which is still a (polyhedral) convex cone [62].

For the case of two settings per party, we have fully characterized the set of NS bilocal correlations: we found 329 different classes of extremal rays, of which 314 are nonlocal. Out of these, 40 are genuinely non-bilocal, i.e., the correlations admit a LHV model but not a bilocal LHV model. A particularly interesting extremal correlation is the following:  $H(A_0, B, C) = H(A_1, B, C) = H(A_1, B) = H(A_0, C) = H(A_1, C) = H(B, C) = 2$  and  $H(A_0, B) = H(A_0) = H(A_1) = H(B) = H(C) = 1$ . It can be understood as the case where Bob and Charlie always measure the same observable (no measurement choice) while Alice still can perform two different measurements. Clearly, since only one of the parties has measurement choices, all correlations arising in this scenario are compatible with a LHV model. However, this correlation is not bilocal, as it can be witnessed by the violation of the entropic inequality valid for any bilocal decomposition:

$$H(A_0, C) \leq H(A_0, B) + H(C|A_1, B). \quad (5)$$

The entropic correlation above arises from a probability

distribution  $p(a, b, c|x) = (1/4)\delta_{a \oplus b, xc}$ , obtained when Alice and Bob share a PR-box  $p_{\text{PR}}$  while Bob and Charlie share a classical correlated distribution  $p_C$ . To that aim, Bob assigns  $b$  to the output of his share of the PR-box that takes as input the bit that is classically correlated with the output  $c$  of Charlie. This result illustrates two novel aspects of the bilocality scenario. First, we see that the nonlocality of the PR-box, which in CHSH scenario is entropically equivalent to a classical correlation, can be activated by employing it in a network. Even more remarkable is the fact that the emergence of nonlocal correlations only requires one out of the three parties to have access to measurement choices. This is similar to what has been observed in [33, 63], where it has been argued that since the role of Charlie can be interpreted as defining measurement choices for Bob, this scenario can be mapped to the CHSH one. In our case, however, we do not need to hinge on this mapping, since we violate a new sort of entropic Bell inequality. Thus, as opposed to Refs. [33, 63], our result does not rely on Bell’s theorem.

Another genuinely nonbilocal extremal entropic ray is associated with the probability  $p(a, b, c|x, y, z) = (1/8)(\delta_{a \oplus b \oplus c, xyz \oplus xy \oplus xz \oplus yz \oplus z \oplus 1} + \delta_{a \oplus b \oplus c, 0})$ . Its nonbilocal can be witnessed via the violation of

$$S_{\text{BL}} = -H_{A_0 B_0 C_0} + H_{A_1 B_1 C_0} + H_{A_0 B_0 C_1} + H_{A_1 B_1 C_1} - H_{A_1, C_1} - H_{A_1, B_1} \geq 0, \quad (6)$$

specifically, with value  $S_{\text{BL}} = -1$ . As opposed to other known inequalities [37, 60, 61, 63], Eq. (6) includes marginal terms, and it is valid for an arbitrary number of outcomes.

*Information causality.*— Information causality (IC) [6] is a principle introduced to explain the limitation of quantum correlations, i.e., Tsirelson bound [64]. It can be understood as a game: Alice receives two independent random bits  $X_0$  and  $X_1$  and the task of Bob is to guess, at each run of the experiment, the value of one of them, having as resources some preshared correlations with Alice and some classical communication ( $H(M)$  bits) sent by her. For shared quantum correlations, the following inequality holds [6]:

$$I(X_0 : G_0) + I(X_1 : G_1) \leq H(M). \quad (7)$$

where  $G_s$  denotes Bob’s guess of  $X_s$ .

To characterize the set of NS entropic correlations associated to IC scenario [i.e, including post-quantum correlations violating (7)], first notice that the mutual information between Alice’s inputs and Bob’s guesses should be limited, according to the assumed causal structure, by the amount of communication, that is,  $I(X_s : G_s) \leq H(M)$ , otherwise they could also communicate superluminally [2]. Here, similarly to what has been done in [6], we consider the marginals  $\mathcal{M}_{\text{IC}} = \{\{X_0, G_0\}, \{X_1, G_1\}, \{M\}\}$ . The NS cone  $\Gamma_{\text{IC}}$  is thus

given by the intersection of the Shannon cone defined by  $\mathcal{M}_{\text{IC}}$  with the constraints  $I(X_y : G_y) \leq H(M)$  ( $y = 0, 1$ ) arising from the causal structure of the game [14]. We found  $\Gamma_{\text{IC}}$  to be characterized by 8 extremal rays, 7 of which respect Eq. (7). The extremal ray violating Eq. (7) is given by  $H(X_0) = H(X_1) = H(G_0) = H(G_1) = H(X_0, G_0) = H(X_1, G_1) = H(M) = 1$ . It is achieved when the parties share a PR-box and apply the protocol used in [6]. It is once more appealing that the NS cone approach naturally retrieves an entropic correlation of special importance.

*Discussion.*— Nonlocality stands nowadays as one of the cornerstones in our understanding of quantum theory. In turn, entropy is a key concept in the foundations and applications of quantum information science. It is thus surprising that still so little is known about their relations and in particular what nonsignaling—another guiding principle permeating all physics—has to say about the entropies that can be generated by the outcomes of physical measurements. Here, we introduced the notion of entropic nonsignaling correlations characterizing the entropies compatible with the fact that we cannot transmit information instantaneously. To illustrate its relevance and novelty, we have applied it to understand a broad range of different phenomena from an entropic perspective: from monogamy relations and nonlocality activation in networks, to genuine multipartite nonlocality.

Nonsignaling also lies at the heart of the device-independent approach to quantum information, which has lately attracted growing attention [31, 65–70], and we believe our results provide a new tool also for practical applications. In addition, the entropic approach provides the natural ground to treat generalized Bell scenarios [58–61] and understand novel forms of nonlocal correlations emerging from it. Future lines of research also include monogamy relations [39], the role of non-Shannon type inequalities [24] in multipartite scenarios and possible applications in nonlocal games [71].

Finally, as demonstrated by information causality [6, 14], many of our current guiding principles are stated in terms of entropy. Our current framework can help to devise new entropic principles, in particular for the multipartite case [72].

## ACKNOWLEDGMENTS

The authors thank Nikolai Miklin for discussions. R.C. acknowledges financial support from the Excellence Initiative of the German Federal and State Governments (Grants ZUK 43 & 81), the FQXi Fund, the US Army Research Office under Contracts W911NF-14-1-0098 and W911NF-14-1-0133 (Quantum Characterization, Verification, and Validation), the DFG (GRO 4334

& SPP 1798). C.B. acknowledges financial support from the EU (Marie Curie CIG 293993/ENFOQI), the FQXi Fund (Silicon Valley Community Foundation), and the DFG.

## I. APPENDIX

### A. The Shannon and Bell entropic cones

Given a collection of  $n$  discrete random variables  $X_1, \dots, X_n$ , we denote by  $[n] = \{1, \dots, n\}$  the set of indices and  $2^{[n]}$  its power set. For every  $S \in 2^{[n]}$ , let  $X_S$  be the vector  $(X_i)_{i \in S}$  and  $H(S) := H(X_S)$  the associated Shannon entropy, given by  $H(X_S) := -\sum_{x_S} p(x_S) \log_2 p(x_S)$ . We can define the vector  $\mathbf{h} = (H(\emptyset), H(X_1), \dots, H(X_1, \dots, X_n)) \in R_n := \mathbb{R}^{2^n}$ . Not every vector  $\mathbf{h} \in R_n$  will correspond to an entropy vector, as, e.g., entropies are nonnegative. The entropy cone is defined as the closure of the region

$$\Gamma_E := \overline{\{h \in R_n \mid h_S = H(S) \text{ for some entropy } H\}}.$$

$\Gamma_E$  is known to be a convex cone but a tight and explicit description is still to be found [24]. However, an outer approximation to the entropic cone is known, the so-called Shannon cone.

The Shannon cone  $\Gamma_{\text{Sh}}$  is a polyhedral closed convex cone, i.e., a subset of  $R_n$  defined by a finite set of linear inequalities, known as basic Shannon-type inequalities, plus a normalization constraint. The first type of inequalities are given by monotonicity conditions, for example,  $H(X_{S \cup \{i\}}) \geq H(X_S)$ , stating that the uncertainty about a set of variables should always be larger than or equal to the uncertainty about any subset of it. The second type is the strong subadditivity condition given by  $I(X_i : X_j | X_S) = H(X_{S \cup i}) + H(X_{S \cup j}) - H(X_{S \cup \{i, j\}}) - H(X_S) \geq 0$  which is equivalent to the positivity of the conditional mutual information. The normalization constraint imposes that  $H(\emptyset) = 0$ .

Given  $n$  variables, all the associated Shannon type inequalities (and thus the Shannon cone) are characterized by the following elemental (non-redundant) inequalities [24]

$$\begin{aligned} H([n] \setminus \{i\}) &\leq H([n]), \\ H(S) + H(S \cup \{i, j\}) &\leq H(S \cup \{i\}) + H(S \cup \{j\}), \\ H(\emptyset) &= 0, \end{aligned} \tag{8}$$

for all  $S \subset [n] \setminus \{i, j\}$ ,  $i \neq j$  and  $i, j \in [n]$ . Thus, the Shannon cone associated with  $n$  variables is described by  $2^{n-2} \binom{n}{2} + n$  inequalities plus one normalization constraint.

Given a marginal scenario  $\mathcal{M}$ , we are interested in the projection of  $\Gamma_{\text{Sh}}$  in the subspace of  $R_n$  representing only observable terms, that is, the Bell cone  $\Gamma_{\text{Bell}}$  associated with the marginal scenario in question. Since we

are given a description of the Shannon cone in terms of linear inequalities, to perform this projection we need to eliminate from this system of inequalities all terms corresponding to non-observables terms. In practice, this is achieved via a Fourier-Motzkin (FM) elimination [73]. The final set of inequalities obtained via the FM elimination (and after eliminating over redundant inequalities) gives all the facets of the associated Bell cone, the non-trivial of which are exactly the entropic Bell inequalities [17, 18, 20, 21]. By non-trivial, we denote those inequalities that are not simply basic Shannon type inequalities like (8).

To exemplify the general framework it is useful to analyze the particular case of the CHSH scenario discussed in the main text. In this case, the existence of an underlying LHV model implies the existence of  $H(A_0, A_1, B_0, B_1)$  (and all its marginals) respecting the elementary inequalities in (8). Proceeding with the FM elimination and keeping only the observable terms  $H(A_i, B_j)$ ,  $H(A_i)$  and  $H(B_j)$  (with  $i, j = 0, 1$ ) we observe that the only non-trivial inequality (up to permutations) is given by

$$\begin{aligned} &H(A_0, B_0) + H(A_0, B_1) + H(A_1, B_0) \\ &- H(A_1, B_1) - H(A_0) - H(B_0) \geq 0. \end{aligned} \quad (9)$$

This is exactly the inequality originally derived in [17]. Replacing the bipartite entropies by mutual informations in (9), that is, using  $H(A_x, B_y) = H(A_x) + H(B_y) - I(A_x : B_y)$ , we obtain the form (2) discussed in the main text and that can be understood as the entropic analogue of the CHSH inequality. However, opposed to the probabilistic version of the CHSH inequality, its entropic version can be used as a nonlocality witness for an arbitrary finite number of measurement outcomes.

### B. A tool for the derivation of entropic Bell inequalities

The formalism outlined above provides a general framework for the derivation of entropic Bell inequalities in basically any scenario [18, 20, 21]. In short, we have to perform the projection –via a Fourier-Motzkin (FM) elimination – of the Shannon cone  $\Gamma_{\text{Sh}}$  to the subspace of observable coordinates defining the Bell cone  $\Gamma_{\text{Bell}}$ . The problem with this approach is that it turns out to be computationally extremely demanding. The FM algorithm eliminate a variable  $x$  from a system of inequalities by summing, after proper normalization, inequalities where  $x$  appears with coefficient +1 and with coefficient -1, and keeping the rest. Eliminating over  $m$  variables in a system of  $N$  inequalities can lead to a number  $O(N^{2^m})$  of inequalities, that is, double exponential. Since the number of initial inequalities describing  $\Gamma_{\text{Sh}}$  is itself exponential [24], this leads to a triple-exponential complexity algorithm, which is

limited, in practice, to very simple cases. In fact, no systematic characterization of entropic inequalities is known beyond particular instances of the bipartite case, although particular multipartite inequalities have been derived [21, 23]. In particular, no entropic Bell inequality witnessing genuine multipartite nonlocality [34] was known to this date.

Nevertheless, checking whether a given inequality is valid for a scenario of interest is computationally much simpler. First, notice that any valid entropic inequality must follow from a FM elimination over the system of inequalities defining the scenario at question, for example, entropic Bell inequalities follow from the basic Shannon type inequalities (8) plus possible additional linear constraints. Given the entropy vector  $\mathbf{h} \in \mathbb{R}_n$ , any linear inequality can be written as the inner product  $\langle \mathcal{I}, \mathbf{h} \rangle \geq 0$ , where  $\mathcal{I}$  is a vector to the inequality. Similarly, a system of inequalities, e.g., Eq. (8), can be represented as a matrix  $M$ , such that  $\Gamma := \{\mathbf{h} | M\mathbf{h} \geq \mathbf{0}\}$ . To check the validity of an inequality  $\langle \mathcal{I}, \mathbf{h} \rangle \geq 0$  with respect to the system  $M\mathbf{h} \geq \mathbf{0}$ , one simply needs to solve the following (efficient) linear program [24]:

$$\begin{aligned} &\text{minimize } \langle \mathcal{I}, \mathbf{h} \rangle \\ &\quad \mathbf{h} \in \mathbb{R}^{2^n} \\ &\text{subject to } M\mathbf{h} \geq \mathbf{0}. \end{aligned} \quad (10)$$

If the minimum of  $\langle \mathcal{I}, \mathbf{h} \rangle$  is larger or equal to zero, then the inequality is valid.

Moreover, one can extend this result to inequalities for the projected cone, even without performing the corresponding FM elimination on the system  $M\mathbf{h} \geq \mathbf{0}$ . More precisely, given  $\Gamma := \{\mathbf{h} | M\mathbf{h} \geq \mathbf{0}\}$  and  $\Pi_k(\Gamma) = \{\mathbf{h} | \tilde{M}\mathbf{h} \geq \mathbf{0}, (\mathbf{h})_l = 0 \text{ for } l > k\}$ , where  $\Pi_k$  is the projection on the first  $k$  coordinates, we prove that the inequality  $\langle \mathcal{I}, \mathbf{h} \rangle \geq 0$ , with  $(\mathcal{I})_l = 0$  for  $l > k$  is valid for the polyhedral cone  $\Pi_k(\Gamma)$  if and only if it is valid for the polyhedral cone  $\Gamma$ . If there exists  $\mathbf{p} \in \Gamma$  such that  $\langle \mathcal{I}, \mathbf{p} \rangle < 0$ , then  $\Pi_k(\mathbf{p}) \in \Pi_k(\Gamma)$  and  $\langle \mathcal{I}, \Pi_k(\mathbf{p}) \rangle = \langle \mathcal{I}, \mathbf{p} \rangle$ . Vice versa, given  $\tilde{\mathbf{p}} \in \Pi_k(\Gamma)$  such that  $\langle \mathcal{I}, \tilde{\mathbf{p}} \rangle < 0$ , there exists  $\mathbf{p} \in \Gamma$  such that  $\Pi_k(\mathbf{p}) = \tilde{\mathbf{p}}$  and  $\langle \mathcal{I}, \mathbf{p} \rangle = \langle \mathcal{I}, \tilde{\mathbf{p}} \rangle$ .

Similarly, it is again a linear program (in fact, a feasibility problem) to check if a given observed entropic correlation  $\tilde{\mathbf{h}}_{\text{obs}}$  is local or nonlocal. It is simply given by

$$\text{minimize } \langle \mathcal{I}, \mathbf{h} \rangle \quad \mathbf{h} \in \mathbb{R}^{2^n} \quad (11)$$

$$\begin{aligned} &\text{subject to } M\mathbf{h} \geq \mathbf{0}, \\ &\quad \mathbf{h}_{\text{obs}} = \tilde{\mathbf{h}}_{\text{obs}}. \end{aligned} \quad (12)$$

where  $\mathcal{I}$  in this case can be any linear objective function. We see that in this case, we not only impose the constraints  $M\mathbf{h} \geq \mathbf{0}$  but also impose that some of the coordinates of the vector  $\mathbf{h}$  (those given by  $\mathbf{h}_{\text{obs}}$ ) should correspond to the observable quantities  $\tilde{\mathbf{h}}_{\text{obs}}$ . In case the



linear program above has no solution (non-feasible), the correlations  $\mathbf{h}_{\text{obs}}$  at question are thus nonlocal.

These linear programs together with the characterization of the entropic NS cone provide novel tools in the derivation of entropic Bell inequalities, as explained below.

A nonlocal extremal correlation given by  $\mathbf{h}_{\text{obs}}^{\text{NS}}$  will imply a non-feasible LP of the form (11). However, not all the constraints encoded in the matrix  $M$  will be necessary for witnessing this non-feasibility: many of the inequalities in  $M$  can be eliminated until we reach a (not unique) minimum set of inequalities that will still lead to a non-feasible LP. Given this minimum set we can then perform a FM elimination that thus will lead to an inequality that is violated by the given correlation  $\mathbf{h}_{\text{obs}}^{\text{NS}}$ . Notice that in general the obtained inequalities are not necessarily going to be facets of the marginal cone of interest. However, we can obtain tight inequalities by adding noise to the extremal correlations (for instance, white noise). By doing that we guarantee that we are deriving inequalities probing the nonlocal character of the given correlation until the noise is strong enough to make it enter in the marginal cone of interest (and thus become local). Notice that a similar procedure can be applied to obtain the minimum set of inequalities required to prove the validity of a given inequality bounding the marginal cone of interest.

### C. Entropic NS cones in a variety of scenarios

For the entropic NS cone description, we use the elemental Shannon type inequalities (8) for each subset of mutually compatible variables defined by a given marginal scenario. In the sections below, we describe in details the entropic NS cone in the bipartite and tripartite scenario, and the hybrid local-nonsignaling models. Notice that for  $n \leq 3$ , the Shannon cone correspond with the true entropy cone, so our bounds are tight. For each of the scenarios we consider, we have catalogued all the inequivalent classes of extremal rays and characterized whether they correspond to local or nonlocal correlations.

#### 1. Bipartite

In a bipartite Bell scenario, two parties Alice and Bob can measure  $m$  possible different observables. Thus, the bipartite entropic NS cone  $\Gamma_{\text{NS}}^2$  is defined by the elemental inequalities for each subset of mutually compatible observables  $A_x, B_y$ , that is, it is simply described in terms of the elemental inequalities  $I(A_x : B_y) \geq 0$ ,  $H(A_x, B_y) \geq H(A_x)$  and  $H(A_x, B_y) \geq H(B_y)$  for  $x, y = 0, \dots, m-1$ .

As discussed in the main text, for the case  $m = 2$  (corresponding to the CHSH scenario)  $\Gamma_{\text{NS}}^2$  is characterized by 5 different classes of extremal NS rays, 4 of which correspond to local correlations and only 1 correspond to nonlocal correlations. In Table (I) we list all the different classes (a class is defined by the inequalities that are equivalent up to the permutation of parties and/or observables).

The class #1 of extremal rays corresponds to the case where one of the variables has maximal entropy, e.g.,  $H(A_0) = \log_2 d$  (where  $d$  is the number of outcomes), while all other variables have null entropy (that is, they represent probability distributions with deterministic outcomes). Class #2 represents perfect (anti)correlations between one observable of Alice and one of Bob, e.g.,  $H(A_0) = H(B_0) = I(A_0 : B_0) = \log_2 d$  while all other entropies are null. Class #3 represents perfect (anti)correlations between the two observables of Alice with one observable of Bob, e.g.,  $H(A_x) = H(B_0) = I(A_x : B_0) = \log_2 d$ , while the other observable of Bob has null entropy. Class #4 represents perfect (anti)correlations between all the observables of Alice and Bob, that is,  $H(A_x) = H(B_y) = I(A_x : B_y) = \log_2 d$ . All these rays clearly correspond to local correlations.

In turn, the class #5 corresponds to entropic nonlocal correlations. For  $d = 2$ , it can be understood as the entropic version of the paradigmatic PR-box [3]

$$p_{\text{PR}}(a, b|x, y) = \begin{cases} 1/2, & a \oplus b = xy \\ 0, & \text{otherwise} \end{cases}. \quad (13)$$

First, notice that  $p_{\text{PR}}$  is entropically equivalent to the purely classical correlations [51]

$$p_{\text{PC}}(a, b|x, y) = \begin{cases} 1/2, & a \oplus b = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (14)$$

That is, both (13) and (14) correspond to the class #4 of extremal rays. The reason for this equivalence is due to the fact that entropies are unable to distinguish between correlations and anti-correlations [51]. Interestingly, the class #5 correspond to a probability distribution obtained by mixing with equal weights the distributions (13) and (14), that is,  $p_{\text{class}\#5} = (1/2)(p_{\text{PR}} + p_{\text{PC}})$  corresponding to three perfect (anti)correlated pairs of variables  $I(A_0 : B_0) = I(A_0 : B_1) = I(A_1 : B_0) = 1$  and one uncorrelated pair  $I(A_1 : B_1) = 0$ . The class #5 violates the entropic version of the CHSH inequality (9) up to the algebraic maximum, achieving  $S_E = \log_2 d$ .

We have performed the same analysis for the case  $m = 3$  and obtained 20 different classes of extremal rays, 7 of which correspond local and 13 to nonlocal correlations. The different classes of extremal rays are listed in Table (II). The interpretation of the local rays is identical to the one we have detailed above to the case of 2 measurement settings. Regarding the nonlocal rays, we focus attention to the classes #8 and #19. Class

Extremal ray									
#	$H(A_x)$		$H(B_y)$		$H(A_x B_y)$				$x / y / xy$
	0	1	0	1	00	01	10	11	
1	1	0	0	0	1	1	0	0	
2	1	0	1	0	1	1	1	1	
3	1	1	1	0	1	1	1	1	
4	1	1	1	1	1	1	1	1	
5	1	1	1	1	1	1	1	2	

TABLE I. Inequivalent classes of extremal rays in the bipartite scenario where both parties measure two possible observables. Extremal rays 1-4 correspond to local entropies while ray 5 correspond to the entropic analog of the PR-box [3]. All rays are defined up to a positive constant multiplicative factor, that is, the rays are described by a vector in some real space  $v \in \mathbb{R}^n$  such that  $\{\lambda v | \lambda \geq 0\}$

#8 basically correspond to the entropic version of the PR-box discussed above, since  $H(A_2) = H(B_2) = 0$ . In turn, class #19 can be understood as the NS correlation that maximally violates the entropic version of the Collins-Gisin inequality [74] given by [21]

$$\begin{aligned} S_3 &= I(A_0 : B_2) - I(A_0 : B_1) + I(A_1 : B_1) \\ &\quad - I(A_1 : B_0) + I(A_1 : B_2) + I(A_2 : B_2) + I(A_2 : B_1) \\ &\quad + I(A_2 : B_0) - H(A_2) - 2H(B_2) - H(B_1) \leq 0, \end{aligned} \quad (15)$$

reaching  $S_3 = 2 \log_2 d$ .

### 2. Tripartite

In a tripartite Bell scenario, three parties Alice, Bob and Charlie perform different measurements in their shares of a joint state. Thus, the tripartite entropic NS cone  $\Gamma_{\text{NS}}^3$  is defined by the elemental inequalities for each subset of mutually compatible observables  $A_x, B_y, C_z$  (with  $x, y, z = 0, \dots, m-1$ ).

We have obtained the full characterization of  $\Gamma_{\text{NS}}^3$  in terms of extremal rays for  $m = 2$ . There are 1292 different inequivalent classes of extremal rays, 128 of which are local and 1164 are nonlocal. Furthermore, as discussed in details below, in the tripartite case we can introduce the notion of genuine tripartite nonlocal (GTNL) correlations. From the 1164 different classes of extremal nonlocal rays, 932 of them are GTNL.

There are thus 232 rays corresponding to nonlocal correlations but displaying no GTNL. We focus our attention to a particular class of these rays, given by  $H(A_x) = H(B_y) = H(C_z) = 1$ ,  $H(A_x, B_y) = H(A_x, C_z) = H(B_y, C_z) = 2$  and  $H(A_0, B_0, C_0) = H(A_1, B_1, C_0) = H(A_1, B_0, C_1) = H(A_1, B_1, C_1) = 2$  and  $H(A_1, B_0, C_0) = H(A_0, B_1, C_0) = H(A_0, B_0, C_1) = H(A_0, B_1, C_1) = 3$ . It can be obtained by the mixing of the nonlocal correlation

$$p(a, b, c | x, y, z) = \begin{cases} 1/2, & a \oplus b \oplus c = yz \oplus x \oplus y \oplus z \\ 0, & \text{otherwise} \end{cases}, \quad (16)$$

with the classical correlation

$$p_{\text{PC}}(a, b, c | x, y, z) = \begin{cases} 1/2, & a \oplus b \oplus c = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (17)$$

The nonlocal character of this distribution can be witnessed via the following tripartite entropic inequality (obtained via the approach described in Sec. IB)

$$\begin{aligned} M_3 &= H_{A_0 B_1 C_1} - H_{A_1 B_1 C_1} - H_{A_1 B_1 C_0} - H_{A_1 B_0 C_1} \\ &\quad - H_{A_0 B_0 C_0} + H_{B_0 C_0} + H_{A_1 C_1} + H_{A_1 B_1} \leq 0, \end{aligned} \quad (18)$$

since the correlations above imply that  $M_3 = 1$  thus violating the inequality. To prove this inequality analytically it is sufficient to consider the chain rule for entropies, implying that

$$\begin{aligned} H_{A_1 B_1 C_1 A_0 B_0 C_0} &= H_{A_0 | B_1 C_1 A_1 B_0 C_0} + H_{B_0 | C_1 A_1 B_1 C_0} \\ &\quad + H_{C_1 | A_1 B_1 C_0} + H_{A_1 | B_1 C_0} + H_{B_1 | C_0} + H_{C_0}. \end{aligned} \quad (19)$$

Using the basic inequalities saying that  $H(A) \leq H(A, B)$  and  $H(A | B, C) \leq H(A | B)$ , we can turn the chain rule decomposition above in the inequality

$$\begin{aligned} H_{A_0 B_1 C_1} &= H_{A_0 | B_0 C_0} + H_{B_0 | C_1 A_1} + H_{C_1 | A_1 B_1} \\ &\quad + H_{A_1 | B_1 C_0} + H_{B_1 | C_0} + H_{C_0}, \end{aligned} \quad (20)$$

that can be rewrite exactly as  $M_3 \leq 0$  if we use that  $H(A | B) = H(A, B) - H(B)$ .

### 3. Proving the monogamy inequality for the entropic CHSH

In order to prove that the monogamy inequality (3) of the main text holds for nonsignaling correlations, we must show that it follows from the elemental inequalities for each subset of mutually compatible observables  $A_x, B_y, C_z$ . It is sufficient to add the the following ele-



Extremal rays																		
# $x / y / xy$	$H(A_x)$			$H(B_y)$			$H(A_x B_y)$											
	0	1	2	0	1	2	00	01	02	10	11	12	20	21	22			
1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0			
2	1	0	0	1	0	0	1	1	1	1	0	0	1	0	0			
3	1	1	0	1	0	0	1	1	1	1	1	1	1	0	0			
4	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0			
5	1	1	1	1	0	0	1	1	1	1	1	1	1	0	1			
6	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1			
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
8	1	1	0	1	1	0	1	1	1	1	2	1	1	1	0			
9	1	1	1	1	1	1	2	2	1	2	2	1	1	1	1			
10	1	1	1	1	1	1	2	1	2	2	2	1	1	1	1			
11	1	1	1	1	1	1	2	1	2	1	2	1	2	1	1			
12	1	1	1	1	1	0	2	1	1	2	1	1	1	1	1			
13	1	1	1	1	1	0	1	2	1	2	1	1	1	1	1			
14	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1			
15	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1			
16	1	1	1	1	1	1	1	1	2	1	2	1	2	1	1			
17	1	1	1	1	1	0	2	1	1	1	1	1	1	1	1			
18	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1			
19	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1			
20	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1			

TABLE II. Inequivalent classes of extremal rays in the bipartite scenario where both parties measure three possible observables. Extremal rays 1-7 correspond to local entropies while rays 8-20 correspond to nonlocal correlations.

mental inequalities

$$H_{A_0 B_0} + H_{A_0 C_1} \geq H_{A_0 B_0 C_1} + H_{A_0}, \quad (21)$$

$$H_{A_0 B_1} + H_{A_0 C_0} \geq H_{A_0 B_1 C_0} + H_{A_0}, \quad (22)$$

$$H_{A_1 B_0} + H_{B_0 C_1} \geq H_{A_1 B_0 C_1} + H_{B_0}, \quad (23)$$

$$H_{A_1 C_0} + H_{B_1 C_0} \geq H_{A_1 B_1 C_0} + H_{C_0}, \quad (24)$$

$$H_{A_0 B_0 C_1} \geq H_{B_0 C_1}, \quad (25)$$

$$H_{A_0 B_1 C_0} \geq H_{B_1 C_0}, \quad (26)$$

$$H_{A_1 B_0 C_1} \geq H_{A_1 C_1}, \quad (27)$$

$$H_{A_1 B_1 C_0} \geq H_{A_1 B_1}, \quad (28)$$

leading to

$$H_{A_0 B_0} + H_{A_0 B_1} + H_{A_1 B_0} - H_{A_1 B_1} - H_{A_0} - H_{B_0} \quad (29)$$

$$+ H_{A_0 C_0} + H_{A_0 C_1} + H_{A_1 C_0} - H_{A_1 C_1} - H_{A_0} - H_{C_0} \geq 0,$$

that can be rewritten exactly as

$$S_E^{AB} + S_E^{AC} \leq 0, \quad (30)$$

with

$$S_E^{AB} = I_{A_0:B_0} + I_{A_0:B_1} + I_{A_1:B_0} - I_{A_1:B_1} - H_{A_0} - H_{B_0}, \quad (31)$$

and

$$S_E^{AC} = I_{A_0:C_0} + I_{A_0:C_1} + I_{A_1:C_0} - I_{A_1:C_1} - H_{A_0} - H_{C_0}. \quad (32)$$

A similar construction can be used to show that the monogamy relation (30) holds for any symmetry of the

entropic CHSH inequalities (31) and (32), thus proving that whenever entropic the marginal correlations  $H(A_x, B_y)$  display nonlocality necessarily  $H(A_x, C_z)$  must be local.

#### 4. Genuine tripartite nonlocality

Similarly to what happens to entanglement [54], when moving from the bipartite case, different classes of nonlocality can be categorized. In the tripartite scenario one can introduce the notion of genuine tripartite nonlocality, that is, a stronger form of nonlocality that cannot be reproduced by LHV models even if two of the parties are allowed to share some nonlocal resources. As discussed in the main text, hybrid local-nonsignaling (L|NS) models are those that can be decomposed as

$$p(a, b, c | x, y, z) = \sum_{\lambda} p(a | x, \lambda) p(b, c | y, z, \lambda) p(\lambda) + \sum_{\mu} p(a, b | x, y, \mu) p(c | z, \mu) p(\mu) + \sum_{\nu} p(a, c | x, z, \nu) p(b | y, \nu) p(\nu). \quad (33)$$

where  $\sum_{\lambda} p(\lambda) + \sum_{\mu} p(\mu) + \sum_{\nu} p(\nu) = 1$  and where  $p(b, c | y, z, \lambda)$  (similarly to the other permutations) represent some nonlocal resource. The different nonlocal resources we allow the parties to share will lead

to distinct notions of genuine multipartite nonlocality. See [35, 36] for a discussion of the different (non-signalling or signalling) nonlocal resources that can be used in the definition of genuine multipartite nonlocality. Here we will focus our attention to nonsignalling nonlocal resources, as for example, nonlocal quantum correlations or PR-boxes [3]. In this case, (33) is equivalent to the existence of probability distributions  $p(a_0, a_1, b_j, c_k)$ ,  $p(a_i, b_0, b_1, c_k)$  and  $p(a_i, b_j, c_0, c_1)$  such that the marginals  $p(a_i, b_j, c_k)$  coincide  $\forall i, j, k = 0, 1$ .

To simplify the discussion let us consider that each of the parties can perform two possible measurements. The first term in the decomposition (33), that is, a model with decomposition given by

$$p(a, b, c|x, y, z) = \sum_{\lambda} p(a|x, \lambda) p^{\text{NS}}(b, c|y, z, \lambda) p(\lambda), \quad (34)$$

is equivalent, in the entropic description, to the existence of  $H(A_0, A_1, B_j, C_k)$  and all its marginals. For each value of  $j, k = 0, 1$  we have therefore a collection of four variables respecting the Shannon type inequalities (8).

Similarly, the two other terms in (33), namely

$$p(a, b, c|x, y, z) = \sum_{\mu} p^{\text{NS}}(a, b|x, y, \mu) p(c|z, \mu) p(\mu), \quad (35)$$

$$p(a, b, c|x, y, z) = \sum_{\nu} p^{\text{NS}}(a, c|x, z, \nu) p(b|y, \nu) p(\nu), \quad (36)$$

imply the existence of  $H(A_i, B_0, B_1, C_k)$  and  $H(A_i, B_j, C_0, C_1)$  respectively. Thus, following the general prescription, the entropic description of hybrid local-nonsignaling (L|NS) correlations corresponds to the intersection of the Shannon cones defined for each of subsets of variables  $\{A_0, A_1, B_j, C_k\}$ ,  $\{A_i, B_0, B_1, C_k\}$  and  $\{A_i, B_j, C_0, C_1\}$  with  $i, j, k = 0, 1$ .

The marginal entropic cone  $\Gamma_{\text{L|NS}}$  characterizing (33) will be the convex hull of the of the cones  $\Gamma_{\text{L|NS}}^a, \Gamma_{\text{L|NS}}^b$  and  $\Gamma_{\text{L|NS}}^c$  characterizing, respectively, (34), (35) and (36). We have tried computationally to compute the extremal rays of  $\Gamma_{\text{L|NS}}$  however the computation seems to be out of reach. In spite of that, checking whether a given nonlocal extremal ray of  $\Gamma_{\text{NS}}^3$  defines or not a genuine tripartite nonlocal correlation is computationally much simpler.

Notice that  $\Gamma_{\text{L|NS}} \subset \Gamma_{\text{NS}}^3$  and since  $\Gamma_{\text{L|NS}}$  is obtained as the convex hull of  $\Gamma_{\text{L|NS}}^i$  ( $i = a, b, c$ ), checking whether a given extremal ray of  $\Gamma_{\text{NS}}^3$  lies inside  $\Gamma_{\text{L|NS}}$  is equivalent to check whether it lies inside each of the cones  $\Gamma_{\text{L|NS}}^i$  ( $i = a, b, c$ ). This can be seen as follows. Let  $r \in \Gamma_{\text{NS}}^3$ , then  $r \in \Gamma_{\text{L|NS}}$  if  $r = h_1 + h_2 + h_3$ , where  $h_i \in \Gamma_{\text{L|NS}}^i$ . Notice that, by cone properties, the sum can be interpreted as a the convex sum  $\frac{1}{3}(3h_1 + 3h_2 + 3h_3)$ . This can be written as a linear program (actually, as feasibility problem, i.e.,  $\mathcal{I}$  is irrelevant)

$$\text{minimize } \langle \mathcal{I}, h \rangle \quad (37)$$

$$\text{subject to } M_1 h_1 \geq 0, \quad (38)$$

$$M_2 h_2 \geq 0,$$

$$M_3 h_3 \geq 0,$$

$$h_1 + h_2 + h_3 = r.$$

In addition, if  $r$  is an extremal ray of  $\Gamma_{\text{NS}}^3$ , then  $r = h_1 + h_2 + h_3$  implies that at least two of the  $h_i$  are zero (extremal rays cannot be written as convex combinations). Hence, for extremal rays it is sufficient to check that they belong to at least one of the  $\Gamma_{\text{L|NS}}^i$ .

Thus, running three sets of linear programs, one for each cone  $\Gamma_{\text{L|NS}}^i$  ( $i = a, b, c$ ), we can decide whether a given extremal entropic NS ray displays or not genuine tripartite nonlocality.

A similar construction holds to check whether a given entropic inequality is a valid witness of GTNL. An entropic inequality  $\mathcal{I}$  for GTNL should be satisfied by any correlation that can be written as the convex sum over points within  $\Gamma_{\text{L|NS}}^i$  ( $i = 1, 2, 3$ ), that is,

$$\langle \mathcal{I}, h_1 + h_2 + h_3 \rangle \geq 0, \quad (39)$$

where  $h_i \in \Gamma_{\text{L|NS}}^i$ . Any inequality satisfying  $\langle \mathcal{I}, h_1 \rangle \geq 0$ ,  $\langle \mathcal{I}, h_2 \rangle \geq 0$  and  $\langle \mathcal{I}, h_3 \rangle \geq 0$  will also satisfy (39). This means that in practice to check if a given inequality is a valid GTNL inequality, we need to solve three linear programs like (10), one for each set of inequalities characterizing the cones  $\Gamma_{\text{L|NS}}^i$  ( $i = 1, 2, 3$ ). This can be seen as follows. Consider the LP

$$\text{minimize } \langle \mathcal{I}, h_1 + h_2 + h_3 \rangle \quad (40)$$

$$\text{subject to } M_1 h_1 \geq 0, \quad (41)$$

$$M_2 h_2 \geq 0,$$

$$M_3 h_3 \geq 0.$$

The inequality is valid if and only if  $\min \langle \mathcal{I}, h_1 + h_2 + h_3 \rangle \geq 0$ . However, since  $h_i = 0$  is a valid solution for all three cones, the above is equivalent to require separately that  $\min \langle \mathcal{I}, h_i \rangle \geq 0$ .

We now turn our attention to the entropic inequality (4) of the main text witnessing GTNL and given by

$$S_{\text{L|NS}} = H_{A_1 B_1 C_0} + H_{A_1 B_0 C_0} + H_{A_1 B_0 C_1} + H_{A_0 B_1 C_0} + H_{A_0 B_1 C_1} - H_{A_1 B_1 C_1} - H_{A_1 B_0} - H_{A_1 C_0} - H_{A_0 C_1} - H_{B_1 C_0} \geq 0. \quad (42)$$

To prove that this inequality is valid for the cone  $\Gamma_{\text{L|NS}}^a$ , it is sufficient to consider the sum of the following basic inequalities (valid for the subsets of variables

$\{A_0, A_1, B_j, C_k\}$ :

$$\begin{aligned}
H_{A_0 A_1 B_1 C_0} &\geq H_{A_0 A_1 C_0}, \\
H_{A_0 A_1 B_1 C_1} &\geq H_{A_1 B_1 C_1}, \\
H_{A_0 A_1 B_0 C_0} &\geq H_{A_0 A_1 B_0}, \\
H_{A_0 A_1 B_0 C_1} &\geq H_{A_0 A_1 C_1}, \\
H_{A_1 B_1 C_0} + H_{A_0 B_1 C_0} &\geq H_{A_0 A_1 B_1 C_0} + H_{B_1 C_0}, \\
H_{A_0 A_1 C_1} + H_{A_0 B_1 C_1} &\geq H_{A_0 A_1 B_1 C_1} + H_{A_0 C_1}, \\
H_{A_0 A_1 C_0} + H_{A_1 B_0 C_0} &\geq H_{A_0 A_1 B_0 C_0} + H_{A_1 C_0}, \\
H_{A_0 A_1 B_0} + H_{A_1 B_0 C_1} &\geq H_{A_0 A_1 B_0 C_1} + H_{A_1 B_0}.
\end{aligned} \tag{43}$$

To prove the validity of (42) for the cones  $\Gamma_{\text{L|NS}}^b$  and  $\Gamma_{\text{L|NS}}^c$  a similar sum of eight basic inequalities is again enough.

As discussed in the main text, the entropic inequality (42) can be used to witness the presence of genuine tripartite nonlocality in quantum states. To that aim we consider  $d$ -dimensional GHZ states

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |jjj\rangle, \tag{44}$$

and projective measurements given by [56]

$$|k\rangle_{p,m} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} e^{i \frac{2\pi}{d} j(k + \alpha_{p,m})} |j\rangle, \tag{45}$$

where  $p = 1, 2, 3$  denotes the party,  $m = 0, 1$  denotes its measurement choice and  $k$  denotes its outcome. For this sort of measurements on GHZ states the probabilities of outcomes can readily be computed to be given by

$$p(a, b, c | x, y, z) = \frac{1}{d^4} \csc^2 \left( \frac{\gamma_{a,b,c,x,y,z}}{d} \right) \sin^2 \left( \gamma_{a,b,c,x,y,z} \right),$$

with  $\gamma_{a,b,c,x,y,z} = \pi(a + b + c + \alpha_{1,x} + \alpha_{2,y} + \alpha_{3,z})$ .

Using this expression we can readily compute the associated Shannon entropies and thus perform numerical optimizations to obtain the optimal violation of inequality (42). We have performed such analysis up to  $d = 40$  and the results are shown in Fig. 2 of the main text.

## 5. Bilocality

The so-called bilocality scenario [37] has been introduced as the classical analogue of an entanglement swapping experiment [57], where two independent pairs of entangled particles are distributed among three parties. Joint measurements on one particle of each pair (e.g. in a Bell basis) can create entanglement and non-local correlations among the remaining two particles, even though they have never interacted. Formally, we

consider that a central node, Bob, receives one particle from both pairs while two other parts, Alice and Charlie, receive each one particle from a given entangled pair. As usual in a Bell scenario, we consider that at each run of the experiment each of the parties can perform different possible measurements labeled by random variables  $X, Y$  and  $Z$  and obtain, respectively, outcomes labeled by  $A, B, C$ . The observed probability distribution  $p(a, b, c | x, y, z)$  can thus be decomposed as

$$p(a, b, c | x, y, z) = \sum_{\lambda_1, \lambda_2} p(\lambda_1) p(\lambda_2) p(a | x, \lambda_1) p(b | y, \lambda_1, \lambda_2) p(c | z, \lambda_2). \tag{46}$$

Notice that in (46) the finer structure of the underlying causal structure is taken into account, namely in the fact that  $p(\lambda_1, \lambda_2) = p(\lambda_1) p(\lambda_2)$  (independence of the sources) and that the outcomes  $A$  of Alice ( $C$  of Charlie) only depend on  $\Lambda_1$  ( $\Lambda_2$ ).

Considering that each of the parties perform two possible measurements, the entropic description of bilocality is equivalent to the existence of a joint entropy  $H(A_0, A_1, B_0, B_1, C_0, C_1)$  respecting the Shannon type inequalities together with the bilocality constraint  $H(A_0, A_1, C_0, C_1) = H(A_0, A_1) + H(C_0, C_1)$ .

We further notice that in the bilocality scenario there are two possible slightly different situations we may want to consider. In the first scenario the random variables  $B_0$  and  $B_1$  stand for the measurement outcomes of two different measurements performed by Bob. In this case  $H(B_0, B_1)$  will correspond to a non-observable quantity. In the second situation, we can understand  $B_0$  and  $B_1$  as standing for a finer description of a single measurement performed by Bob. For instance, consider that Bob always measure in a Bell basis  $\{|\Psi^\pm\rangle, |\Phi^\pm\rangle\}$ . The variable  $B_0$  could stand for the information regarding whether the measurement outcome correspond to  $|\Psi\rangle$  or  $|\Phi\rangle$  while  $B_1$  would stand for the information about phase, for instance,  $|\Psi^+\rangle$  or  $|\Psi^-\rangle$ . In this situation,  $H(B_0, B_1)$  corresponds to an observable quantity. In the following we will focus our attention to the first case, that is, we have to eliminate from our description terms like  $H(B_0, B_1)$ .

Similarly to what happens in the usual tripartite case, the complete description of the bilocality scenario in terms of a FM elimination is out of computational reach. However, characterizing the extremal rays of the bilocal entropic NS cone is much simpler. We have obtained 329 different classes of extremal NS bilocal entropic rays, 15 of which are bilocal while 314 correspond to nonbilocal correlations. From this 314 rays, 40 are genuine-nonbilocal in the sense of admitting a LHV model but not a bilocal LHV model. Using the tool described in Sec. IB one can derive entropic inequalities detecting the nonbilocality of these rays. Next we employ the framework in Sec. IB to provide an analytical

proof for the bilocality inequalities (5) and (6) discussed in the main text. To prove (5), we have to sum the following basic inequalities

$$\begin{aligned} H_{A_0} + H_C &\geq H_{A_0 C}, \\ H_{A_0 A_1 B C} &\geq H_{A_0 A_1 C}, \\ H_{A_0 A_1} + H_{A_0 B} &\geq H_{A_0 A_1 B} + H_{A_0}, \\ H_{A_0 A_1 B} + H_{A_1 B C} &\geq H_{A_0 A_1 B C} + H_{A_1 B}. \end{aligned} \quad (47)$$

and use the bilocality constraint  $H_{A_0 A_1} + H_C = H_{A_0 A_1 C}$ .

In turn, to prove inequality (6) we have to sum the following Shannon type inequalities

$$\begin{aligned} H_{A_0 A_1 B_0 B_1 C_0 C_1} &\geq H_{A_0 A_1 B_0 B_1 C_0}, \\ H_{A_0 A_1 B_0 B_1 C_0 C_1} &\geq H_{A_0 A_1 B_0 C_0 C_1}, \\ H_{B_0 C_1} + H_{C_0 C_1} &\geq H_{B_0 C_0 C_1} + H_{C_1}, \\ H_{A_0 A_1} + H_{A_1 C_1} &\geq H_{A_0 A_1 C_1} + H_{A_1}, \\ H_{A_0 B_0 C_1} + H_{B_0 C_0 C_1} &\geq H_{A_0 B_0 C_0 C_1} + H_{B_0 C_1}, \\ H_{A_1 B_1 C_0} + H_{A_1 B_1 C_1} &\geq H_{A_1 B_1 C_0 C_1} + H_{A_1 B_1}, \\ H_{A_0 A_1 C_1} + H_{A_1 C_0 C_1} &\geq H_{A_0 A_1 C_0 C_1} + H_{A_1 C_1}, \\ H_{A_0 B_0 B_1 C_0} + H_{A_0 B_0 C_0 C_1} &\geq H_{A_0 B_0 B_1 C_0 C_1} + H_{A_0 B_0 C_0}, \\ H_{A_1 B_0 C_0 C_1} + H_{A_1 B_1 C_0 C_1} &\geq H_{A_1 B_0 B_1 C_0 C_1} + H_{A_1 C_0 C_1}, \\ H_{A_0 A_1 B_0 B_1 C_0} + H_{A_0 B_0 B_1 C_0 C_1} &\geq H_{A_0 A_1 B_0 B_1 C_0 C_1} + H_{A_0 B_0 B_1 C_0}, \\ H_{A_0 A_1 B_0 C_0 C_1} + H_{A_1 B_0 B_1 C_0 C_1} &\geq H_{A_0 A_1 B_0 B_1 C_0 C_1} + H_{A_1 B_0 C_0 C_1}. \end{aligned} \quad (48)$$

and use the bilocality constraint  $H_{A_0 A_1} + H_{C_0 C_1} = H_{A_0 A_1 C_0 C_1}$ .

## 6. Information Causality

Information causality (IC) [6] has been introduced as a principle to explain the degree of nonlocality of quantum mechanics. It basically states that the amount of information a given part, Bob can have over some bits in possession of another part, Alice, is limited by the amount of information communicated from Alice to Bob. Information causality is respected by quantum correlations, however, as shown in [6] any nonlocal correlation stronger than maximum permitted by quantum mechanics, in sense of surpassing the Tsirelson's bound [64] for the CHSH inequality, will violate it.

IC can be understood as a game between Alice and Bob: Alice receives two independent random bits  $X_0$  and  $X_1$  and the aim of Bob at each run of the game is to guess the value of one of them, having as resources some pre-shared correlations with Alice and a certain amount of communication sent by her. Which of the two bits he should guess is decided by a random variable  $S$ . In a classical description we have the variables  $X_0, X_1$  representing the input bits of Alice; the variables  $Y_0, Y_1$  standing for the guesses of Bob ( $Y_s$  is the guess of bit  $X_s$  given that  $S = s$ );  $M$  stands for the message sent

from Alice to Bob and  $\Lambda$  corresponds to the pre-shared correlations between them.

The (classical) entropic description of IC is equivalent to the existence of a joint entropy  $H(X_0, X_1, B_0, B_1, M, \Lambda)$  fulfilling the following causal constraints imposed by the rules of the game:  $I(X_0 : X_1) = 0$ ,  $I(X_0, X_1 : \Lambda) = 0$  and  $I(X_0, X_1 : B_0, B_1 | M, \Lambda) = 0$ . The first constraint encodes the fact that both bits are uncorrelated and thus have null mutual information (though this is not strictly necessary [14, 75]). The second constraint encodes the fact that the bits received by Alice are independent of the pre-shared correlation with Bob. Finally, the third constraint encodes the fact that conditioned on the message sent by Alice and on the pre-shared correlations, the guesses of Bob should become completely uncorrelated with the input bits. A similar description can be given for the case where the pre-shared correlations arise from an entangled state  $\rho$  rather than a classical variable  $\Lambda$  [14].

Proceeding with the FM elimination we obtain the description of the marginal entropic cone characterizing the information causality game. To that aim, first one needs to define the marginal scenario of interest. Here, similarly to what has been done in [6], we will focus on the marginal scenario  $\mathcal{M}_{IC} = \{\{X_0, Y_0\}, \{X_1, Y_1\}, \{M\}\}$ . Notice, however, that more general marginal scenarios can be defined [14]. It follows that the only non-trivial entropic inequality describing the IC game (for classical and quantum pre-shared correlations) is given by:

$$I(X_0 : Y_0) + I(X_1 : Y_1) \leq H(M). \quad (49)$$

This inequality quantifies the qualitative expectation that the amount of information Bob has about Alice's inputs is bounded by the amount of information contained in the message (as quantified by  $H(M)$ ).

We are now in position to define the set of entropic NS correlation for the IC scenario. For the marginal scenario under consideration, the NS set is simply characterized by the Shannon type inequalities  $I(X_s : Y_s) \geq 0$ ,  $H(X_s, Y_s) \geq H(X_s)$ ,  $H(X_s, Y_s) \geq H(Y_s)$ ,  $H(M) \geq 0$  together with the constraint  $I(X_s : Y_s) \leq H(M)$ . The NS cone is described in terms of 8 extremal rays (shown in the Table (IC 5)), 7 corresponding to correlations that respect the information causality constraint (rays 1 to 7) and only 1 violating it (ray 8). The extremal ray violating (49) is exactly given by the entropic correlations obtained if we replace the pre-shared correlation by a PR-box and apply the protocol discussed in the original IC paper [6]: Alice inputs  $x = x_0 \oplus x_1$  in her part of the PR-box, obtaining an outcome  $a$  that is then used to encode the message  $m = x_0 \oplus a$ . Bob inputs in his part of the PR-box,  $y = 0, 1$  depending on which bit  $X_y$  he is interested and obtains an outcome  $b_y$ . His guess is given

Extremal ray	$H(X_0)$	$H(X_1)$	$H(Y_0)$	$H(Y_1)$	$H(X_0, Y_0)$	$H(X_1, Y_1)$	$H(M)$
1	0	0	0	0	0	0	1
2	0	0	0	1	0	1	0
3	0	0	1	0	1	0	0
4	0	1	0	0	0	1	0
5	1	0	0	0	1	0	0
6	0	1	0	1	0	1	1
7	1	0	1	0	1	0	1
8	1	1	1	1	1	1	1

TABLE III. All the extremal rays defining the entropic NS cone of the information causality scenario.

by  $g_y = m \oplus b_y = x_0 \oplus a \oplus b_y = x_0 \oplus y(x_0 \oplus x_1)$ , that

is, he can perfectly guess both outputs of Alice thought only one bit of information ( $H(M) = 1$ ) has been sent.

- 
- [1] J. S. Bell, *Physics* **1**, 195 (1964).  
[2] S. Popescu, *Nat. Phys.* **10**, 264 (2014).  
[3] S. Popescu and D. Rohrlich, *Foundations of Physics* **24**, 379 (1994).  
[4] M. Navascués and H. Wunderlich, *Proc. Roy. Soc. Lond. A* **466**, 881 (2010).  
[5] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, *Nat. Commun.* **6**, 6288 (2015).  
[6] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, *Nature* **461**, 1101 (2009).  
[7] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, *Nat. Commun.* **4**, 2263 (2013).  
[8] A. Cabello, *Phys. Rev. Lett.* **110**, 060402 (2013).  
[9] A. B. Sainz, T. Fritz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, *Phys. Rev. A* **89**, 032117 (2014).  
[10] W. van Dam, *Nonlocality & communication complexity*, Ph.D. thesis, Faculty of Physical Sciences, University of Oxford (1999).  
[11] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, *Phys. Rev. Lett.* **96**, 250401 (2006).  
[12] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Phys. Rev. A* **81**, 062348 (2010).  
[13] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, *Phys. Rev. Lett.* **104**, 230404 (2010).  
[14] R. Chaves, C. Majenz, and D. Gross, *Nat. Commun.* **6**, 5766 (2015).  
[15] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).  
[16] S. Abramsky and A. Brandenburger, *New Journal of Physics* **13**, 113036 (2011).  
[17] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **61**, 662 (1988).  
[18] R. Chaves and T. Fritz, *Phys. Rev. A* **85**, 032113 (2012).  
[19] N. J. Cerf and C. Adami, *Phys. Rev. A* **55**, 3371 (1997).  
[20] T. Fritz and R. Chaves, *IEEE Trans. Inform. Theory* **59**, 803 (2013).  
[21] R. Chaves, L. Luft, and D. Gross, *New Journal of Physics* **16**, 043001 (2014).  
[22] P. Kurzyński and D. Kaszlikowski, *Phys. Rev. A* **89**, 012103 (2014).  
[23] S. Raeisi, P. Kurzyński, and D. Kaszlikowski, *Phys. Rev. Lett.* **114**, 200401 (2015).  
[24] R. W. Yeung, *Information theory and network coding*, Information technology–transmission, processing, and storage (Springer, 2008).  
[25] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, *New Journal of Physics* **12**, 033024 (2010).  
[26] O. C. O. Dahlsten, D. Lercher, and R. Renner, *New Journal of Physics* **14**, 063024 (2012).  
[27] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, and B. Schölkopf, *The Annals of Statistics* **41**, 2324 (2013).  
[28] R. Chaves, L. Luft, T. O. Maciel, D. Gross, D. Janzing, and B. Schölkopf, *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 112 (2014).  
[29] J. Henson, R. Lal, and M. F. Pusey, *New J. Phys.* **16**, 113043 (2014).  
[30] H. S. Poh, M. Markiewicz, P. Kurzyński, A. Cerè, D. Kaszlikowski, and C. Kurtsiefer, *arXiv preprint arXiv:1504.03126* (2015).  
[31] R. Chaves, J. B. Brask, and N. Brunner, *Phys. Rev. Lett.* **115**, 110501 (2015).  
[32] D. Janzing, R. Chaves, and B. Schoelkopf, *arXiv preprint arXiv:1512.02057* (2015).  
[33] T. Fritz, *New J. Phys.* **14**, 103001 (2012).  
[34] G. Svetlichny, *Phys. Rev. D* **35**, 3066 (1987).  
[35] R. Gallego, L. E. Wurfli, A. Acín, and M. Navascués, *Phys. Rev. Lett.* **109**, 070401 (2012).  
[36] J.-D. Bancal, J. Barrett, N. Gisin, and S. Pironio, *Phys. Rev. A* **88**, 014102 (2013).  
[37] C. Branciard, N. Gisin, and S. Pironio, *Phys. Rev. Lett.* **104**, 170401 (2010).  
[38] L. Masanes, A. Acín, and N. Gisin, *Phys. Rev. A* **73**, 012112 (2006).  
[39] M. Pawłowski and C. Brukner, *Phys. Rev. Lett.* **102**, 030403 (2009).  
[40]  $\mathbf{p}_{\text{Obs}}$  denotes the vector of probabilities with components given by all possible values of  $a_x$  and  $b_y$ .  
[41] A. Fine, *Phys. Rev. Lett.* **48**, 291 (1982).  
[42] I. Pitowsky, *Quantum probability–quantum logic*, Lecture notes in physics (Springer-Verlag, 1989).

- [43] I. Pitowsky, *Mathematical Programming* **50**, 395 (1991).
- [44] (), this convex cone structure arises from the from the possibility of making copies of two or more random variables, e.g.,  $\tilde{\mathbf{h}} = \mathbf{h} + \mathbf{h}'$  if  $\tilde{X} = (X, X')$ , or making copies conditioning on a value of a third variable, e.g.,  $X$  is either 0,  $Y$ , or  $Z$  depending on the value of a third variable  $U$ . See Ref. [24] for more details.
- [45] C. D. Aliprantis and R. Tourky, *Cones and duality* (American Mathematical Soc., 2007).
- [46] C. Budroni and A. Cabello, *Journal of Physics A: Mathematical and Theoretical* **45**, 385304 (2012).
- [47] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2009).
- [48] (), this implies, in particular, that for  $n \leq 3$  every NS entropic vector has a realization in terms of NS probability distribution.
- [49] See Supplemental Material.
- [50] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- [51] R. Chaves, *Phys. Rev. A* **87**, 022102 (2013).
- [52] S. Pironio, J.-D. Bancal, and V. Scarani, *Journal of Physics A: Mathematical and Theoretical* **44**, 065303 (2011).
- [53] (), these cones are polyhedral because of the polyhedral structure of the entropy cone for two and three variables.
- [54] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [55] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, *Phys. Rev. A* **71**, 022101 (2005).
- [56] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, *Phys. Rev. Lett.* **88**, 040404 (2002).
- [57] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, *Phys. Rev. Lett.* **71**, 4287 (1993).
- [58] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, *Phys. Rev. A* **90**, 062109 (2014).
- [59] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, *Phys. Rev. Lett.* **114**, 140403 (2015).
- [60] R. Chaves, *Phys. Rev. Lett.* **116**, 010402 (2016).
- [61] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, *Phys. Rev. Lett.* **116**, 010403 (2016).
- [62] (), this is no longer true for affine subspace, which are never used in this context.
- [63] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, *Phys. Rev. A* **85**, 032119 (2012).
- [64] B. S. Cirel'son, *Letters in Mathematical Physics* **4**, 93 (1980).
- [65] A. K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991).
- [66] J. Barrett, L. Hardy, and A. Kent, *Phys. Rev. Lett.* **95**, 010503 (2005).
- [67] R. Colbeck, arXiv preprint arXiv:0911.3814 (2009).
- [68] S. Pironio, A. Acín, S. Massar, A. B. de La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, *et al.*, *Nature* **464**, 1021 (2010).
- [69] R. Colbeck and R. Renner, *Nat. Phys.* **8**, 450 (2012).
- [70] R. Gallego, N. Brunner, C. Hadley, and A. Acín, *Phys. Rev. Lett.* **105**, 230501 (2010).
- [71] C. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, *Phys. Rev. Lett.* **92**, 127901 (2004).
- [72] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, *Phys. Rev. Lett.* **107**, 210403 (2011).
- [73] H. P. Williams, *American Mathematical Monthly*, 681 (1986).
- [74] D. Collins and N. Gisin, *Journal of Physics A: Mathematical and General* **37**, 1775 (2004).
- [75] S. W. Al-Safi and A. J. Short, *Physical Review A* **84**, 042323 (2011).